DOCUMENT RESUME

ED 258 312 CS 504 971

AUTHOR Pisoni, David B.; And Others
TITLE Speech Perception, Word Recognition and the Structure of the Lexicon. Research on Speech Perception Progress Report No. 10.
INSTITUTION Indiana Univ., Bloomington.
SPONS AGENCY National Institutes of Health (DHHS), Bethesda, Md.
PUB DATE May 85
GRANT NIH-NS-12179-08
NOTE 31p.; Paper presented at the Annual Meeting of the Midwestern Psychological Association (Chicago, IL, May 2-4, 1985).
PUB TYPE Reports - Research/Technical (143) -- Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Auditory Perception; Communication Research; *Dictionaries; *Learning Theories; *Listening Comprehension; Models; Phonemes; Simulation; Speech Communication; *Word Recognition
IDENTIFIERS *Cohort Theory of Word Recognition; Phonetic Refinement Theory; *Theory Development

ABSTRACT

The results of three projects concerned with auditory word recognition and the structure of the lexicon are reported in this paper. The first project described was designed to test experimentally several specific predictions derived from MACS, a simulation model of the Cohort Theory of word recognition. The second project description provides the results of analyses of the structure and distribution of words in the lexicon using a large lexical database. In this discussion, statistics about similarity spaces for high and low frequency words are applied to previously published data on the intelligibility of words presented in noise, and differences in identification are shown to be related to structural factors about the specific words and the distribution of similar words in their neighborhoods. Finally, the third project description reports efforts at developing a new theory of word recognition known as the Phonetic Refinement Theory, which was designed to incorporate some of the detailed acoustic-phonetic and phonotactic knowledge that listeners have about the internal structure of words and the organization of words in the lexicon, and about how they use this knowledge in word recognition. (Author/HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED258312

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 10 (1984) Indiana University]

Speech Perception, Word Recognition and the Structure of the Lexicon*

David B. Pisoni, Howard C. Nusbaum, Paul A. Luce,

and Louisa M. Slowiaczek

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

David B. Pisoni

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

*Preparation of this paper was supported, in part, by NIH research grant NS-12179-08 to Indiana University in Bloomington. We thank Beth Greene for her help in editing the manuscript, Arthur House and Tom Crystal for providing us with a machine readable version of the lexical database used in our analyses and Chris Davis for his outstanding contributions to the software development efforts on the SRL Project Lexicon. This paper was written in honor of Ludmilla Chistovich, one of the great pioneers of speech research, on her 60th birthday. We hope that the research described in this paper will influence other researchers in the future in the way Dr. Chistovich's now classic work has influenced our own thinking about the many important problems in speech perception and production. It is an honor for us to submit this report as a small token of our appreciation of her contributions to the field of speech communications. We extend our warmest regards and congratulations to her on this occasion and wish her many more years of good health and productivity. This paper will appear in Speech Communication, 1985.

CS 504971

Abstract

This paper reports the results of three projects concerned with auditory word recognition and the structure of the lexicon. The first project was designed to experimentally test several specific predictions derived from MACS, a simulation model of the Cohort Theory of word recognition. Using a priming paradigm, evidence was obtained for acoustic-phonetic activation in word recognition in three experiments. The second project describes the results of analyses of the structure and distribution of words in the lexicon using a large lexical database. Statistics about similarity spaces for high and low frequency words were applied to previously published data on the intelligibility of words presented in noise. Differences in identification were shown to be related to structural factors about the specific words and the distribution of similar words in their neighborhoods. Finally, the third project describes efforts at developing a new theory of word recognition known as Phonetic Refinement Theory. The theory is based on findings from human listeners and was designed to incorporate some of the detailed acoustic-phonetic and phonotactic knowledge that human listeners have about the internal structure of words and the organization of words in the lexicon, and how they use this knowledge in word recognition. Taken together, the results of these projects demonstrate a number of new and important findings about the relation between speech perception and auditory word recognition, two areas of research that have traditionally been approached from quite different perspectives in the past.

Speech Perception, Word Recognition and the Structure of the Lexicon

Introduction

Much of the research conducted in our laboratory over the last few years has been concerned, in one way or another, with the relation between early sensory input and the perception of meaningful linguistic stimuli such as words and sentences. Our interest has been with the interface between the acoustic-phonetic input -- the physical correlates of speech -- on the one hand, and more abstract levels of linguistic analysis that are used to comprehend the message. Research on speech perception over the last thirty years has been concerned principally, if not exclusively, with feature and phoneme perception in isolated CV or CVC nonsense syllables. This research strategy has undoubtedly been pursued because of the difficulties encountered when one deals with the complex issues surrounding the role of early sensory input in word recognition and spoken language understanding and its interface with higher levels of linguistic analysis. Researchers in any field of scientific investigation typically work on tractable problems and issues that can be studied with existing methodologies and paradigms. However, relative to the bulk of speech perception research on isolated phoneme perception, very little is currently known about how the early sensory-based acoustic-phonetic information is used by the human speech processing system in word recognition, sentence perception or comprehension of fluent connected speech.

Several general operating principles have guided the choice of problems we have decided to study. We believe that continued experimental and theoretical work is needed in speech perception in order to develop new models and theories that can capture significant aspects of the process of speech sound perception and spoken language understanding. To say, as some investigators have, that speech perception is a "special" process requiring specialized mechanisms for perceptual analysis is, in our view, only to define one of several general problems in the field of speech perception and not to provide a principled explanatory account of any observed phenomena. In our view, it is important to direct research efforts in speech perception toward somewhat broader issues that use meaningful stimuli in tasks requiring the use of several sources of linguistic knowledge by the listener.

Word Recognition and Lexical Representation in Speech

Although the problems of word recognition and the nature of lexical representations have been long-standing concerns of cognitive psychologists, these problems have not generally been studied by investigators working in the mainstream of speech perception research (see [1,2]). For many years these two lines of research, speech perception and word recognition, have remained more-or-less distinct from each other. This was true for several reasons. First, the bulk of work on word recognition was concerned with investigating visual word recognition processes with little, if any, attention directed to questions of spoken word recognition. Second, most of the interest and research effort in speech perception was directed toward feature and phoneme perception. Such an approach is appropriate for studying the "low level" auditory analysis of speech but it is not useful in dealing with questions

surrounding how words are recognized in isolation or in connected speech or how various sources of knowledge are used by the listener to recover the talker's intended message.

Many interesting and potentially important problems in speech perception involve the processes of word recognition and lexical access and bear directly on the nature of the various types of representations in the mental lexicon. For example, at the present time, it is of considerable interest to determine precisely what kinds of representations exist in the mental lexicon. Do words, morphemes, phonemes, or sequences of spectral templates characterize the representation of lexical entries? Is a word accessed on the basis of an acoustic, phonetic or phonological code? Why are high frequency words recognized so rapidly? We are interested in how human listeners hypothesize words for a given stretch of speech. Furthermore, we are interested in characterizing the sensory information in the speech signal that listeners use to perceive words and how this information interacts with other sources of higher-level linguistic knowledge. These are a few of the problems we have begun to study in our laboratory over the past few years.

Past theoretical work in speech perception has not been very well developed, nor has the link between theory and empirical data been very sophisticated. Moreover, work in the field of speech perception has tended to be defined by specific experimental paradigms or particular phenomena (see [3,4]). The major theoretical issues in speech perception often seem to be ignored, or alternatively, they take on only a secondary role and therefore receive little serious attention by investigators who are content with working on the details of specific experimental paradigms.

Over the last few years, some work has been carried out on questions surrounding the interaction of knowledge sources in speech perception, particularly research on word recognition in fluent speech. A number of interesting and important findings have been reported recently in the literature and several models of spoken word recognition have been proposed to account for a variety of phenomena in the area. In the first section of this paper we will briefly summarize several recent accounts of spoken word recognition and outline the general assumptions that follow from this work that are relevant to our own recent research. Then we will identify what we see as the major issues in word recognition. Finally, we will summarize the results of three ongoing projects that use a number of different research strategies and experimental paradigms to study word recognition and the structure of the lexicon. These sections are designed to give the reader an overview of the kinds of problems we are currently studying as we attempt to link research in speech perception with auditory word recognition.

Word Recognition and Lexical Access

Before proceeding, it will be useful to distinguish between word recognition and lexical access, two terms that are often used interchangeably in the literature. We will use the term word recognition to refer to those computational processes by which a listener identifies the acoustic-phonetic and/or phonological form of spoken words (see [5]). According to this view, word recognition may be simply thought of as a form of pattern recognition. The sensory and perceptual processes used in word recognition are assumed to

be the same whether the input consists of words or pronounceable nonwords. We view the "primary recognition process" as the problem of characterizing how the form of a spoken utterance is recognized from an analysis of the acoustic waveform. This description of word recognition should be contrasted with the term lexical access which we use to refer to those higher-level computational processes that are involved in the activation of the meaning or meanings of words that are currently present in the listener's mental lexicon (see [5]). By this view, the meaning of a word is accessed from the lexicon after its phonetic and/or phonological form makes contact with some appropriate representation previously stored in memory.

Models of Word Recognition

A number of contemporary models of word recognition have been concerned with questions of processing words in fluent speech and have examined several types of interactions between bottom-up and top-down sources of knowledge. However, little, if any, attention has been directed at specifying the precise nature of the early sensory-based input or how it is actually used in word recognition processes. Klatt's recent work on the LAFS model (Lexical Access From Spectra) is one exception [6]. His proposed model of word recognition is based on sequences of spectral templates in networks that characterize the properties of the sensory input. One important aspect of Klatt's model is that it explicitly avoids any need to compute a distinct level of representation corresponding to discrete phonemes. Instead, LAFS uses a precompiled acoustically-based lexicon of all possible words in a network of diphone power spectra. These spectral templates are assumed to be context-sensitive like "Wickelphones" [7] because they characterize the acoustic correlates of phones in different phonetic environments. They accomplish this by encoding the spectral characteristics of the segments themselves and the transitions from the middle of one segment to the middle of the next.

Klatt [6] argues that diphone concatenation is sufficient to capture much of the context-dependent variability observed for phonetic segments in spoken words. According to this model, word recognition involves computing a spectrum of the input speech every 10 ms and then comparing this input spectral sequence with spectral templates stored in the network. The basic idea, adopted from HARPY, is to find the path through the network that best represents the observed input spectra [8]. This single path is then assumed to represent the optimal phonetic transcription of the input signal.

Another central problem in word recognition and lexical access deals with the interaction of sensory input and higher-level contextual information. Some investigators, such as Forster [9,10] and Swinney [11] maintain that early sensory information is processed independently of higher-order context, and that the facilitation effects observed in word recognition are due to post-perceptual processes involving decision criteria (see also [12]). Other investigators such as Morton [13,14,15], Marslen-Wilson and Tyler [16], Tyler and Marslen-Wilson [17,18], Marslen-Wilson and Welsh [19], Cole and Jakimik [20] and Foss and Blank [21] argue that context can, in fact, influence the extent of early sensory analysis of the input signal.

Although Foss and Blank [21] explicitly assume that phonemes are computed during the perception of fluent speech and are subsequently used during the process of word recognition and lexical access, other investigators such as Marslen-Wilson and Welsh [19] and Cole and Jakimik [20,22] have argued that words, rather than phonemes, define the locus of interaction between the initial sensory input and contextual constraints made available from higher sources of knowledge. Morton's [13,14,15] well-known Logogen Theory of word recognition is much too vague, not only about the precise role that phonemes play in word recognition, but also as to the specific nature of the low-level sensory information that is input to the system.

It is interesting to note in this connection that Klatt [6], Marslen-Wilson and Tyler [16] and Cole & Jakimik [22] all tacitly assume that words are constructed out of linear sequences of smaller elements such as phonemes. Klatt implicitly bases his spectral templates on differences that can be defined at a level corresponding to phonemes: likewise, Marslen-Wilson and Cole & Jakimik implicitly differentiate lexical items on the basis of information about the constituent segmental structure of words. This observation is, of course, not surprising since it is precisely the ordering and arrangement of different phonemes in spoken languages that specifies the differences between different words. The ordering and arrangement of phonemes in words not only indicates where words are different but also how they are different from each other (see [23] for a brief review of these arguments). These relations therefore provide the criterial information about the internal structure of words and their constituent morphemes required to access the meanings of words from the lexicon.

Although Klatt [6] argues that word recognition can take place without having to compute phonemes along the way, Marslen-Wilson has simply ignored the issue entirely by placing his major emphasis on the lexical level. According to his view, top-down and bottom-up sources of information about a word's identity are integrated together to produce what he calls the primary recognition decision which is assumed to be the immediate lexical interpretation of the input signal. Since Marslen-Wilson's "Cohort Theory" of word recognition has been worked out in some detail, and since it occupies a prominent position in contemporary work on auditory word recognition and spoken language processing, it will be useful to summarize several of the assumptions and some of the relevant details of this approach. Before proceeding to Cohort Theory, we examine several assumptions of its predecessor, Morton's Logogen Theory.

Logogen and Cohort Theory of Word Recognition

In some sense, Logogen Theory and Cohort Theory are very similar. According to Logogen Theory, word recognition occurs when the activation of a single lexical entry (i.e., a logogen) crosses some critical threshold value [14]. Each word in the mental lexicon is assumed to have a logogen, a theoretical entity that contains a specification of the word's defining characteristics (i.e., its syntactic, semantic, and sound properties). Logogens function as "counting devices" that accept input from both the bottom-up sensory analyzers and the top-down contextual mechanisms. An important aspect of Morton's Logogen Model is that both sensory and contextual information interact in such a way that there is a trade-off relationship

between them; the more contextual information input to a logogen from top-down sources, the less sensory information is needed to bring the Logogen above threshold for activation. This feature of the Logogen model enables it to account for the observed facilitation effects of syntactic and semantic constraints on speed of lexical access (see e.g., [24,25,26]) as well as the word frequency and word apprehension effects reported in the literature. In the presence of constraining prior contexts, the time needed to activate a logogen from the onset of the relevant sensory information will be less than when such constraints are not available because less sensory information will be necessary to bring the logogen above its threshold value.

In contrast to Logogen Theory which assumes activation of only a single lexical item after its threshold value is reached, Cohort Theory views word recognition as a process of eliminating possible candidates by deactivation (see [16,27,28,29,30]). A set of potential word-candidates is activated during the earliest phases of the word recognition process solely on the basis of bottom-up sensory information. According to Marslen-Wilson and Welsh [19], the set of word-initial cohorts consists of the entire set of words in the language that begins with a particular initial sound sequence. The length of the initial sequence defining the initial cohort is not very large, corresponding roughly to the information in the first 200-250 ms of a word. According to the Cohort Theory, a word is recognized at the point that a particular word can be uniquely distinguished from any of the other words in the word-initial cohort set that was defined exclusively by the bottom-up information in the signal. This is known as the "critical recognition point" of a word. Upon first hearing a word, all words sharing the same initial sound characteristics become activated in the system. As the system detects mismatches between the initial bottom-up sensory information and the top-down information about the expected sound representation of words generated by context, inappropriate candidates within the initial cohort are deactivated.

In Cohort Theory, as in the earlier Logogen Theory, word recognition and subsequent lexical access are viewed as a result of a balance between the available sensory and contextual information about a word at any given time. In particular, when deactivation occurs on the basis of contextual mismatches, less sensory information is therefore needed for a single word candidate to emerge. According to the Cohort Theory, once word recognition has occurred the perceptual system carries out a much less detailed analysis of the sound structure of the remaining input. As Marslen-Wilson and Welsh [19] have put it, "No more and no less bottom-up information needs to be extracted than is necessary in a given context", Pp. 58.

Acoustic-Phonetic Priming and Cohort Theory

As outlined above, Cohort theory proposes that in the initial stage of word recognition, a "cohort" of all lexical elements whose words begin with a particular acoustic-phonetic sequence will be activated. Several recent studies in our laboratory (see [31]) have been concerned with testing the extent to which initial acoustic-phonetic information is used to activate a cohort of possible word candidates in word recognition. Specifically, a series of auditory word recognition experiments were conducted using a priming paradigm.

Much of the past research that has used priming techniques was concerned with the influence of the meaning of a prime word on access to the meaning of a target word (e.g., [32]). However, it has been suggested by a number of researchers that the acoustic-phonetic representation of a prime stimulus may also facilitate or inhibit recognition of a subsequent test word (see [33]). A lexical activation model of Cohort Theory called MACS was developed in our lab to test the major assumptions of Cohort Theory [29,31]. Several predictions of the MACS model suggested that phonetic overlap between two items could influence auditory word recognition. Specifically, it was suggested that the residual activation of word candidates following recognition of a prime word could influence the activation of lexical candidates during recognition of a test word. Furthermore, the relationship between the amount of acoustic-phonetic overlap and the amount of residual activation suggested that identification should improve with increasing amounts of acoustic-phonetic overlap between the beginnings of the prime and test words.

In order to test these predictions, we performed an experiment in which subjects heard a prime word followed by a test word. On some trials, the prime and test words were either unrelated or identical. On other trials, although the prime and test words were different, they contained the same initial acoustic-phonetic information. For these trials, the prime and test words shared the same initial phoneme, the first two phonemes or the first three phonemes. Thus, we examined five levels of acoustic-phonetic overlap between the prime and target: 0, 1, 2, 3, or 4 phonemes in common.

By way of example, consider in this context the effects of presenting a single four phoneme word (e.g., the prime) on the recognition system. Following recognition of the prime, the different cohorts activated by the prime will retain a residual amount of activation corresponding to the point at which the candidates were eliminated. When the test word is presented, the effect of this residual activation will depend on the acoustic-phonetic overlap or similarity between the prime and the test word. A prime that shares only the first phoneme of a test word should have less of an effect on identification than a prime that is identical to the test word. The residual activation of the candidates therefore differentially contributes to the rate of reactivation of the cohorts for the test word.

In this experiment, we examined the effect of word primes on the identification of word targets presented in masking noise at various signal-to-noise ratios. Primes and targets were related as outlined above. The prime items were presented over headphones in the clear; targets were presented 50 msec after the prime items embedded in noise. Subjects were instructed to listen to the pair of items presented on each trial and to respond by identifying the second item (the target word embedded in noise). The results of the first experiment supported the predictions of the MACS model and provided support for Cohort Theory. The major findings are shown in Figure 1.

Insert Figure 1 about here

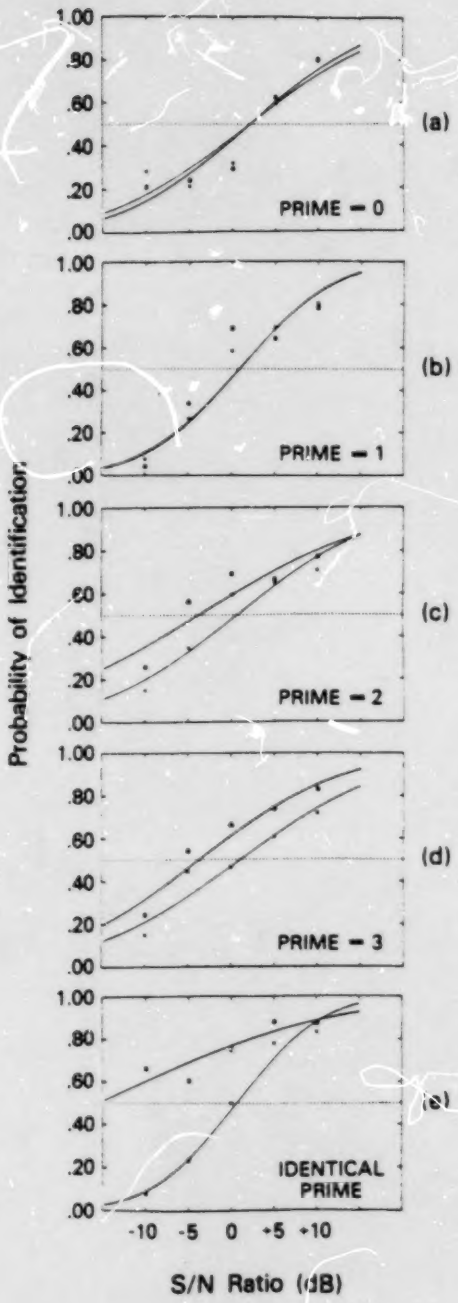

Figure 1. Results displaying the probability of correct identification in a priming experiment using word primes and word targets presented at various signal-to-noise ratios. Crosses in each panel represent unprimed trials and squares represent primed trials when prime-target overlap equals: (a) 0 phonemes, (b) 1 phoneme, (c) 2 phonemes, (d) 3 phonemes and (4) identical (Data from [31]).

Specifically, the probability of correctly identifying targets increased as the acoustic-phonetic overlap between the prime and the target increased. Subjects showed the highest performance in identifying targets when they were preceded by an identical prime. Moreover, probability of correct identification was greater for primes and targets that shared three phonemes than those that shared two phonemes, which were, in turn, greater than pairs that shared one phoneme or pairs that were unrelated.

The results of this experiment demonstrate that acoustic-phonetic priming can be obtained for identification of words that have initial phonetic information in common. However, this experiment did not test the lexical status of the prime. The priming results may have been due to the fact that only word primes preceded the target items. In order to demonstrate that priming was, in fact, based on acoustic-phonetic similarity (as opposed to some lexical effect), we conducted a second identification experiment in which the prime items were phonologically admissable pseudowords. As in the first experiment, the primes shared 3, 2, or 1 initial phonemes with the target or they were unrelated to the target. Because of the difference in lexical status between primes and targets, there was no identical prime-target condition in this experiment. The subject's task was the same as in the first experiment.

As in the first experiment, we found an increased probability of correctly identifying target items as acoustic-phonetic overlap between the prime and target increased. Thus, the lexical status of the prime item did not influence identification of the target. Taken together, the results of both studies demonstrate acoustic-phonetic priming in word recognition. The facilitation we observed in identification of target words embedded in noise suggests the presence of residual activation of the phonetic forms of words in the lexicon. Furthermore, the results provide additional support for the MACS lexical activation model based on Cohort Theory by demonstrating that priming is due to the segment-by-segment activation of lexical representations in word recognition.

One of the major assumptions of Cohort Theory that was incorporated in our lexical activation model is that a set of candidates is activated based on word initial acoustic-phonetic information. Although we obtained strong support for acoustic-phonetic activation of word candidates, the outcome of both experiments did not establish that the acoustic-phonetic information needs to be exclusively restricted to word initial position. In order to test this assumption, we conducted a third priming experiment using the same identification paradigm. In this experiment, word primes and word targets were selected so that the acoustic-phonetic overlap occurred between word primes and targets at the ends of the words. Primes and targets were identical or 0, 1, 2, or 3 phonemes were the same from the end of the words.

As in the first two experiments, we found evidence of acoustic-phonetic priming. The probability of correctly identifying a target increased as the acoustic-phonetic overlap between the prime and target increased from the ends of the items. These results demonstrate that listeners are as sensitive to acoustic-phonetic overlap at the ends of words as they are to overlap at the beginnings of words. According to the MACS model and Cohort Theory, only words that share the initial sound sequences of a prime item should be activated by the prime. Thus, both MACS and Cohort Theory predict that no priming should have been observed. However, the results of the third

experiment demonstrated priming from the ends of words, an outcome that is clearly inconsistent with the predictions of the MACS model and Cohort Theory.

The studies reported here were an initial step in specifying how words might be recognized in the lexicon. The results of these studies demonstrate the presence of some form of residual activation based on acoustic-phonetic properties of words. Using a priming task, we observed changes in word identification performance as a function of the acoustic-phonetic similarity of prime and target items. However, at least one of the major assumptions made about word recognition in Cohort Theory appears to be incorrect. In addition to finding acoustic-phonetic priming from the beginning of words, we also observed priming from the ends of words as well. This latter result suggests that activation of potential word candidates may not be restricted to only a cohort of words sharing word initial acoustic-phonetic information. Indeed, other parts of words may also be used by listeners in word recognition. Obviously, these findings will need to be incorporated into any theory of auditory word recognition. Phonetic Refinement Theory, as outlined in the last section of this paper, was designed to deal with this finding as well as several other problems with Cohort Theory.

Measures of Lexical Density and the Structure of the Lexicon

A seriously neglected topic in word recognition and lexical access has been the precise structural organization of entries in the mental lexicon. Although search theories of word recognition such as Forster's [9,10] have assumed that lexical items are arranged according to word frequency, little work has been devoted to determining what other factors might figure into the organization of the lexicon (see however [34]). Landauer and Streeter [35] have shown that one must take the phonemic, graphemic, and syllabic structure of lexical items into account when considering the word frequency effect in visual recognition experiments. They have shown that a number of important structural differences between common and rare words may affect word recognition. Their results suggest that the frequency and organization of constituent phonemes and graphemes in a word maybe an important determinant of its ease of recognition. Moreover, Landauer and Streeter, as well as Eukel [36], have argued that "similarity neighborhoods" or "phonotactic density" may affect word recognition and lexical access in ways that a simple "experienced" word frequency account necessarily ignores. For example, it would be of great theoretical and practical interest to determine if word recognition is controlled by the relative density of the neighborhood from which a given word is drawn, the frequency of the neighboring items, and the interaction of these variables with the frequency of the word in question. In short, one may ask how lexical distance in this space (as measured, for example, by the Greenberg and Jenkins [37] method) interacts with word frequency in word recognition.

As a first step toward approaching these important issues, we have acquired several large databases. One of these, based on Kenyon and Knott's A Pronouncing Dictionary of American English [38] and Webster's Seventh Collegiate Dictionary [39], contains approximately 300,000 entries. Another smaller database of 20,000 words is based on Webster's Pocket Dictionary. Each entry contains the standard orthography of a word, a phonetic transcription, and special codes indicating the syntactic functions of the word. We have developed a number of algorithms for determining, in various

ways, the similarity neighborhoods, or "lexical density," for any given entry in the dictionary. This information has provided some useful information about the structural properties of words in the lexicon and how this information might be used by human listeners in word recognition.

Lexical Density, Similarity Spaces and the Structure of the Lexicon

Word frequency effects obtained in perceptual and memory research have typically been explained in terms of frequency of usage (e.g., [13,9]), the time between the current and last encounter with the word in question [40], and similar such ideas. In each of these explanations of word frequency effects, however, it has been at least implicitly assumed that high and low frequency words are "perceptually equivalent" [41,42,43,13,44,45]. That is, it has often been assumed that common and rare words are structurally equivalent in terms of phonemic and orthographic composition. Landauer and Streeter [35] have shown, however, that the assumption of perceptual equivalence of high and low frequency words is not necessarily warranted. In their study, Landauer and Streeter demonstrated that common and rare words differ on two structural dimensions. For printed words, they found that the "similarity neighborhoods" of common and rare words differ in both size and composition: High frequency words have more words in common (in terms of one letter substitutions) than low frequency words, and high frequency words tend to have high frequency neighbors, whereas low frequency words tend to have low frequency neighbors. Thus, for printed words, the similarity neighborhoods for high and low frequency words show marked differences. Landauer and Streeter also demonstrated that for spoken words, certain phonemes are more prevalent in high frequency words than in low frequency words and vice versa (see also [46]).

One of us [47] has undertaken a project that is aimed at extending and elaborating the original Landauer and Streeter study (see also [48]). In this research, both the similarity neighborhoods and phonemic constituencies of high and low frequency words have been examined in order to determine the extent to which spoken common and rare words differ in the nature and number of "neighbors" as well as phonemic configuration. To address these issues, an on-line version of Webster's Pocket Dictionary (WPD) was employed to compute statistics about the structural organization of words. Specifically, the phonetic representations of approximately 20,000 words were used to compute similarity neighborhoods and examine phoneme distributions. (See Luce [47] for a more detailed description). Some initial results of this project are reported below.

Similarity Neighborhoods of Spoken Common and Rare Words

In an intial attempt to characterize the similarity neighborhoods of common and rare words, a subset of high and low frequency target words were selected from the WPD for evaluation. High frequency words were defined as those equal to or exceeding 100 words per million in the Kucera and Francis [49] word count. Low frequency words were defined as those between 10 and 30 words per million inclusively. For each target word meeting these a priori frequency criteria, similarity neighborhoods were computed based on

one-phoneme substitutions at each position within the target word. There were 92 high frequency words and 2063 low frequency words. The mean number of words within the similarity neighborhoods for the high and low frequency words were computed, as well as the mean frequencies of the neighbors. In addition, a decision rule was computed as a measure of the distinctiveness of a given target word relative to its neighborhood according to the following formula:

$$\frac{T}{T + \sum N_i};$$

where T equals the frequency of the target word and N equals the frequency of the i-th neighbor of that target word (see [35]). Larger values for the decision rule indicate a target word that "stands out" in its neighborhood; smaller values indicate a target word that is relatively less distinctive in its neighborhood.

Insert Table I about here

The results of this analysis, broken down by the length of the target words, are shown in Table I. (Mean frequencies of less than one were obtained because some words included in the WPD were not listed in Kucera and Francis; these words were assigned a value of zero in the present analysis.) Of primary interest are the data for words of lengths two through four (in which more than two words were found for each length at each frequency). For these word lengths, it was found that although the mean number of neighbors for high and low frequency target words were approximately equal, the mean frequencies of the similarity neighborhoods for high frequency target words of lengths two and three were higher than the mean frequencies of the similarity neighborhoods of the low frequency target words.

No such difference was obtained, however, for target words consisting of four phonemes. Thus, these results only partially replicate Landauer and Streeter's earlier results obtained from printed high and low frequency words, with the exception that the number of neighbors was not substantially different for high and low frequency words nor were the mean frequencies of the neighborhoods different for words consisting of four phonemes.

The finding that high frequency words tend to have neighbors of higher frequency than low frequency words suggests, somewhat paradoxically, that high frequency words are more, rather than less, likely to be confused with other words than low frequency words. At first glance, this finding would appear to contradict the results of many studies demonstrating that high frequency words are recognized more easily than low frequency words. However, as shown in Table I, the decision rule applied to high and low frequency target words predicts that high frequency words should be perceptually distinctive relative to the words in their neighborhoods whereas low frequency targets will not. This is shown by the substantially larger values of this index for high

TABLE I. Similarity neighborhood statistics for high and low frequency words as a function of word length. (Data from [47]).

LENGTH	#WORDS		#NEIGHBORS		MEAN FREQUENCY OF NEIGHBORS	
	HIGH	LOW	HIGH	LOW	HIGH	LOW
1	2	3	24.00	42.00	809.69	560.18
2	36	41	31.39	31.85	527.39	416.95
3	39	278	19.97	22.64	501.22	119.29
4	14	411	6.21	7.88	69.91	69.35
5	1	355	0.00	1.96	1.00	35.66
6	---	277	-----	0.59	-----	14.75
7	---	230	-----	0.33	-----	20.44
8	---	183	-----	0.23	-----	13.63
9	---	148	-----	0.18	-----	7.39
10	---	92	-----	0.12	-----	13.36
11	---	31	-----	0.03	-----	1.00
12	---	13	-----	0.08	-----	5.00
13	---	1	-----	0.00	-----	1.00

LENGTH	DECISION RULE		#UNIQUE		%UNIQUE	
	HIGH	LOW	HIGH	LOW	HIGH	LOW
1	.4635	.0007	0	0	0.00	0.00
2	.2771	.0040	0	0	0.00	0.00
3	.4430	.0357	0	0	0.00	0.00
4	.7962	.1757	1	13	7.14	3.16
5	-----	.4541	---	100	----	28.17
6	-----	.6479	---	174	----	62.82
7	-----	.6340	---	174	----	73.48
8	-----	.7270	---	145	----	79.23
9	-----	.7447	---	126	----	85.14
10	-----	.7377	---	81	----	88.04
11	-----	.9286	---	30	----	96.77
12	-----	.8148	---	12	----	92.31
13	-----	-----	---	1	----	100.00

frequency words than low frequency words of the same length. Work is currently underway in our laboratory to determine if this decision rule predicts identification responses when frequencies of the target words are fixed and the values of the decision rule vary. If the relationship of a target word to its neighborhood, and not the frequency of the target word itself, is the primary predictor of identification performance, this would provide strong evidence that structural factors, rather than experienced frequency per se, underlie the word frequency effect (see also [36,35] for similar arguments).

Also of interest in Table I are the values of the decision rule and the percentage of unique target words (i.e., words with no neighbors) as a function of word length. For target words of both frequencies, the decision rule predicts increasingly better performance for words of greater length (except for the unique situation of one-phoneme high frequency words). In addition, it can be seen that for words consisting of more than three phonemes, the percentage of unique words increases substantially as word length increases. This finding demonstrates that simply increasing the length of a word increases the probability that the phonotactic configuration of that word will be unique and eventually diverge from all other words in the lexicon. Such a result suggests the potentially powerful contribution of word length in combination with various structural factors to the isolation of a given target word in the lexicon.

Phoneme Distributions in Common and Rare Words

The finding that high frequency spoken words tend to be more similar to other high frequency words than to low frequency words also suggests that certain phonemes or phonotactic configurations may be more common in high frequency words than in low frequency words [50,46]. As a first attempt to evaluate this claim, Luce [47] has examined the distribution of phonemes in words having frequencies of 100 or greater and words having a frequency of one. For each of the 45 phonemes used in the transcriptions contained in the WPD, percentages of the total number of possible phonemes for four and five phoneme words were computed for the high and low frequency subsets. (For the purposes of this analysis, function words were excluded. Luce [47] has demonstrated that function words are structurally quite different from content words of equivalent frequencies. In particular, function words tend to have many fewer neighbors than content words. Thus, in order to eliminate any contribution of word class effects, only content words were examined.)

Of the trends uncovered by these analyses, two were the most compelling. First, the percentages of bilabials, interdentals, palatals, and labiodentals tended to remain constant or decrease slightly from the low to high frequency words. However, the pattern of results for the alveolars and velars was quite different. For the alveolars, increases from low to high frequency words of 9.07% for the four phoneme words and 3.63% for the five phoneme words were observed. For the velars, however, the percentage of phonemes dropped from the low to high frequency words by 2.33% and 1.14% for the four and five phoneme words, respectively. In the second trend of interest, there was an increase of 4.84% for the nasals from low to high frequency words accompanied by a corresponding drop of 4.38% in the overall percentage of stops for the five phoneme words.

The finding that high frequency words tend to favor consonants having an alveolar place of articulation and disfavor those having a velar place of articulation suggests that frequently used words may have succumbed to pressures over the history of the language to exploit consonants that are in some sense easier to articulate [50,51]. This result, in conjunction with the finding for five phoneme words regarding the differential use of nasals and stops in common and rare words, strongly suggests that, at least in terms of phonemic constituency, common words differ structurally from rare words in terms of their choice or selection of constituent elements. Further analyses of the phonotactic configuration of high and low frequency words should reveal even more striking structural differences between high and low frequency words in light of the results obtained from the crude measure of structural differences based on the overall distributions of phonemes in common and rare words (see [47]).

Similarity Neighborhoods and Word Identification. In addition to the work summarized above demonstrating differences in structural characteristics of common and rare words, Luce [47] has demonstrated that the notion of similarity neighborhoods or lexical density may be used to derive predictions regarding word intelligibility that surpasses a simple frequency of usage explanation. A subset of 300 words published by Hood and Poole [52] which were ranked according to their intelligibility in white noise has been examined. As Hood and Poole pointed out, frequency of usage was not consistently correlated with word intelligibility scores for their data. It is therefore likely that some metric based on the similarity neighborhoods of these words would be better at capturing the observed differences in intelligibility than simple frequency of occurrence.

To test this possibility, Luce [47] examined 50 of the words provided by Hood and Poole, 25 of which constituted the easiest words and 25 of which constituted the most difficult in their data. In keeping with Hood and Poole's observation regarding word frequency, Luce found that the 25 easiest and 25 most difficult words were not, in fact, significantly different in frequency. However, it was found that the relationship of the easy words to their neighbors differed substantially from the relationship of the difficult words to their neighbors. More specifically, on the average, 56.41% of the words in the neighborhoods of the difficult words were equal to or higher in frequency than the difficult words themselves, whereas only 23.62% of the neighbors of the easy words were of equal or higher frequency. Thus, it appears that the observed differences in intelligibility may have been due, at least in part, to the frequency composition of the neighborhoods of the easy and difficult words, and were not primarily due to the frequencies of the words themselves (see also [53,54]). In particular, it appears that the difficult words in Hoode and Poole's study were more difficult to perceive because they had relatively more "competition" from their neighbors than the easy words.

In summary, the results obtained thus far by Luce suggest that the processes involved in word recognition may be highly contingent on structural factors related to the organization of words in the lexicon and the relation of words to other phonetically similar words in surrounding neighborhoods in the lexicon. In particular, the present findings suggest that the classic word frequency effect may be due, in whole or in part, to structural

differences between high and low frequency words, and not to experienced frequency per se. The outcome of this work should prove quite useful in discovering not only the underlying structure of the mental lexicon, but also in detailing the implications these structural constraints may have for the real-time processing of spoken language by human listeners as well as machines. In the case of machine recognition, these findings may provide a principled way to develop new distance metrics based on acoustic-phonetic similarity of words in large vocabularies.

Phonetic Refinement Theory

Within the last few years three major findings have emerged from a variety of experiments on spoken word recognition (see [22,21,27,19]). First, spoken words appear to be recognized from left-to-right; that is, words are recognized in the same temporal sequence by which they are produced. Second, the beginnings of words appear to be far more important for directing the recognition process than either the middles or the ends of words. Finally, word recognition involves an interaction between bottom-up pattern processing and top-down expectations derived from context and linguistic knowledge.

Although Cohort Theory was proposed to account for word recognition as an interactive process that depends on the beginnings of words for word candidate selection, it is still very similar to other theories of word recognition. Almost all of the current models of human auditory word recognition are based on pattern matching techniques. In these models, the correct recognition of a word depends on the exact match of an acoustic property or linguistic unit (e.g., a phoneme) derived from a stimulus word with a mental representation of that property or unit in the lexicon of the listener. For example, in Cohort Theory, words are recognized by a sequential match between input and lexical representations. However, despite the linear, serial nature of the matching process, most theories of word recognition generally have had little to say about the specific nature of the units that are being matched or the internal structure of words (see [5]). In addition, these theories make few, if any, claims about the structure or organization of words in the lexicon. This is unfortunate because models dealing with the process of word recognition may not be independent from the representations of words or the organization of words in the lexicon.

Recently, two of us [55,56] have proposed a different approach to word recognition that can account for the same findings as Cohort Theory. Moreover, the approach explicitly incorporates information about the internal structure of words and the organization of words in the lexicon. This theoretical perspective, which we have called Phonetic Refinement Theory, proposes that word recognition should be viewed not as pattern matching but instead as constraint satisfaction. In other words, rather than assume that word recognition is a linear process of comparing elements of a stimulus pattern to patterns in the mental lexicon, word recognition is viewed from this perspective as a process more akin to relaxation labeling (e.g., [57]) in which a global interpretation of a visual pattern results from the simultaneous interaction of a number of local constraints. Translating this approach into terms more appropriate for auditory word recognition, the process of identifying a spoken word therefore depends on finding a word in the lexicon that simultaneously satisfies a number of constraints imposed by

the stimulus, the structure of words in the lexicon, and the context in which the word was spoken.

Constraint Satisfaction. Phonetic Refinement Theory is based on the general finding that human listeners can and do use fine phonetic information in the speech waveform and use this information to recognize words, even when the acoustic-phonetic input is incomplete or only partially specified, or when it contains errors or is noisy. At present, the two constraints we consider most important for the bottom-up recognition of words (i.e., excluding the role of lingustic context) are the phonetic refinement of each segment in a word and its word length in terms of the number of segments in the word. Phonetic refinement refers to the process of identifying the phonetic information that is encoded in the acoustic pattern of a word. We assume that this process occurs over time such that each segment is first characterized by an acoustic event description. As more and more acoustic information is processed, acoustic events are characterized using increasingly finer and finer phonetic descriptions. The most salient phonetic properties of a segment are first described (e.g., manner); less salient properties are identified later as more acoustic information accumulates. Thus, we assume that decoding the phonetic structure of a word from the speech waveform requires time during which new acoustic segments are acquired and contribute to the phonetic refinement of earlier segments.

The constraints on word recognition can therefore be summarized as an increasingly better characterization of each of the phonetic segments of a word over time, as well as the development of an overall phonotactic pattern that emerges from the sequence of phonetic segments. These two constraints increase simultaneously over time and can be thought of as narrowing down the set of possible words. At some point, the left-to-right phonotactic constraint converges on the constraint provided by the increasing phonetic refinement of phonetic segments to specify a single word from among a number of phonetically-similar potential candidates.

Organization of the Lexicon. According to this view, words are recognized using a one-pass, left-to-right strategy with no backtracking as in Cohort Theory, LAFS, and Logogen Theory. However, unlike these theories, Phonetic Refinement Theory assumes that words in the lexicon are organized as sequences of phonetic segments in a multi-dimensional acoustic-phonetic space [45]. In this space, words that are more similar in their acoustic-phonetic structures are closer to each other in the lexicon. Furthermore, it is possible to envision the lexicon as structured so that those portions of words that are similar in location and structure are closer together in this space. For example, words that rhyme with each other are topologically deformed to bring together those parts of the words that are phonetically similar and separate those portions of words that are phonetically distinct.

We assume that the recognition process takes place in this acoustic-phonetic space by activating pathways corresponding to words in the lexicon. Partial or incomplete phonetic descriptions of the input activate regions of the lexicon that consist of phonetically similar pathways. As more information is obtained about an utterance by continued phonetic refinement and acquisition of new segments, a progressive narrowing occurs in both the phonetic specification of the stimulus and the set of activated word candidates that are phonetically similar to the input signal. As more segments are acquired from the input and earlier segments are progressively

refined, the constraints on the region of activation are increased until the word is recognized. According to Phonetic Refinement Theory, a word is recognized when the activation path for one word through the phonetic space is higher than any competing paths or regions through the lexicon.

Comparison with Cohort Theory. By focusing on the structural properties of words and the process of constraint satisfaction, Phonetic Refinement Theory is able to account for much of the same data that Cohort Theory was developed to deal with. Moreover, it is able to deal with some of the problems that Cohort Theory has been unable to resolve. First, by allowing linguistic context to serve as another source of constraint on word recognition, Phonetic Refinement Theory provides an interactive account of context effects that is similar to the account suggested by Cohort Theory (cf. [16,19]).

Second, Phonetic Refinement Theory can account for the apparent importance of word beginnings in recognition. In Cohort Theory, the acoustic-phonetic information at the beginning of a word entirely determines the set of cohorts (potential word candidates) that are considered for recognition. Word beginnings are important for recognition by fiat; that is, they are important because the theory has axiomatically assumed that they have a privileged status in determining which candidates are activated in recognition. In contrast, the importance of word beginnings in Phonetic Refinement Theory is simply a consequence of the temporal structure of spoken language and the process of phonetic refinement. Word beginnings do not exclude inconsistent word candidates; rather they activate candidates that are consistent with them to the degree that the candidates are consistent with word-initial information in the signal. Since the beginnings of words are, by necessity, processed first, they receive the most phonetic refinement earliest on in processing and therefore provide the strongest initial constraint on word candidates. As a consequence, Phonetic Refinement Theory can account for the ability of listeners to identify words from only partial information at the beginnings of words (e.g., [27,28,58]). In addition, Phonetic Refinement Theory predicts the finding that subjects can detect nonwords at the first phoneme that causes an utterance to become a nonword; that is, at the point where the nonword becomes different from all the words in the lexicon [59]. The theory makes this prediction directly because the segment that causes the input pattern to become a nonword directs the activation pathway to an "empty" region in the acoustic-phonetic lexical space.

More importantly, Phonetic Refinement Theory can account for a number of results that are inconsistent with Cohort Theory. For example, Cohort Theory cannot directly account for the ability of subjects to identify words in a gating study based on word endings alone [60,28,58]. However, in Phonetic Refinement Theory, word endings are a valid form of constraint on the recognition process and the theory predicts that listeners can and do use this information. The extent to which listeners use word endings in recognition depends, of course, on the relative efficiency of this constraint compared to the constraint provided by word beginnings. Therefore, Phonetic Refinement Theory predicts that listeners should be sensitive to phonetic overlap between prime and test words, whether that overlap occurs at the beginning or the ending of a word (see above and [31] for further details).

In addition, Cohort Theory cannot account for word frequency effects in perception (cf. [15]). The theory incorporates no mechanisms that would predict any effect of frequency whatsoever on word recognition. By comparison, Phonetic Refinement Theory incorporates two possible sources of word frequency effects. The first is based on findings suggesting the possibility that word frequency effects may be explained by the different structural properties of high and low frequency words [36,35]. According to this view, high and low frequency words occupy different acoustic-phonetic regions of the lexicon (see above and Luce [7] for further details). Thus, the density characteristics of these regions in the lexicon could account for the relative ease of perception of high and low frequency words, if high frequency words were in sparse neighborhoods while low frequency words resided in denser regions of the lexicon.

A second account of word frequency effects in perception appeals to the use of experienced frequency or familiarity as a selectional constraint to be used for generating a response once a region of the lexicon has been activated. In the case of isolated words, this selectional constraint represents the subject's "best guess" in the case that no other stimulus properties could be employed to resolve a word path from an activated region. In understanding fluent speech, this selection constraint would be supplanted by the more reasonable constraint imposed by expectations derived from linguistic context [61,62]. Thus, word frequency effects should be substantially attenuated or even eliminated when words are placed in meaningful contexts. This is precisely the result observed by Luce [62] in a study on auditory word identification in isolation and in sentence context.

Finally, Phonetic Refinement Theory is able to deal with the effects of noise, segmental ambiguity, and mispronunciations in a much more elegant manner than Cohort Theory. In Cohort Theory, word-initial acoustic-phonetic information determines the set of possible word candidates from which the recognized word is chosen. If there is a mispronunciation of the initial segment of a word (see [12]), the wrong set of word candidates will be activated and there will be no way to recover gracefully from the error. However, in Phonetic Refinement Theory, if a phonetic segment is incorrectly recognized, two outcomes are possible. If the mispronunciation yields an utterance that is a nonword, correct recognition should be possible by increasing other constraints such as acquiring more segments from the input. At some point in the utterance, the pathway with the highest activation will lead into a "hole" in the lexicon where no word is found. However, the next highest pathway will specify the correct word. An incorrect phoneme that occurs early in a word will probably terminate a pathway in empty space quite early so that by the end of the utterance, the correct word will actually have a higher aggregate level of pathway activation than the aborted path corresponding to the nonword. This is a simple consequence of the correct pathway having more similar segments to the utterance over the entire path than the nonword sequence ending in a hole in the lexicon. If the error occurs late in the word, it may actually occur after the constraints on the word were sufficient to permit recognition. Thus, Phonetic Refinement Theory has little difficulty recovering from errors that result in nonwords. However, for the second type of error -- those that result in a real word other than the intended word -- there is no way the Phonetic Refinement Theory could recover from this error without using linguistic context as a constraint. Of course, this sort of error could not be recovered from by any recognition system, including a human listener, unless context was allowed to

play a direct role in the early recognition process; this assumption is still the topic of intense controversy and we will not attempt to deal with it here.

Structural Constraints on Word Recognition. Although it has been asserted by several researchers that words can be recognized from only a partial specification of the phonetic content of words, these claims are based primarily on data from gating experiments (e.g., [27,60,58,28]). Since we argue that it is the structure of words in the lexicon that determines the performance of human listeners and not simply some form of sophisticated guessing strategy, it is important to learn more about the relative power of different phonetic and phonotactic constraints in reducing the search space of word candidates during recognition.

The approach we have taken to this problem was motivated by several recent studies that were conducted to investigate the relative heuristic power of various classification schemes for large vocabulary word recognition by computers [63,64,34,65]. The goal of this research has been to find a classification scheme that reduces the lexical search space from a very large vocabulary (i.e., greater than 20,000 words) to a very few candidates. An optimal classification heuristic would be one that yields candidate sets that contain an average of one word each, without requiring complete identification of all the phonemes in each word. However, even if one heuristic is not optimal, it may still reduce the search space by a significant amount at a very low cost in computational complexity; other constraints can then be applied to finally "recognize" the word from among the members contained in the reduced search space. Thus, instead of a serial search through a very large number of words, heuristics that reduce the search space can quickly rule out very large subsets of words that are totally inconsistent with an utterance without requiring highly detailed pattern matching.

In a number of recent papers, Zue and his colleagues [64,34] have shown that a partial phonetic specification of every phoneme in a word results in an average candidate set size of about 2 words for a vocabulary of 20,000 words. The partial phonetic specification consisted of six gross manner classes of phonemes. Instead of using 40 to 50 phonemes to transcribe a spoken word, only six gross categories were used: stop consonant, strong fricative, weak fricative, nasal, liquid/glide, or vowel. These categories obviously represent a relatively coarse level of phonetic description and yet when combined with word length, they provide a powerful phonotactic constraint on the size of the lexical search space.

Using a slightly different approach, Crystal et al. [63] demonstrated that increasing the phonetic refinement of every phoneme in a word from four gross categories to ten slightly more refined categories produced large improvements in the number of unique words that could be isolated in a large corpus of text. However, both of these computational studies examined the consequences of partially classifying every segment in a word. Thus, they actually employed two constraints: (1) the partial classification of each segment and (2) the broad phonotactic shape of each word resulting from the combination of word length with gross phonetic category information.

Insert Figure 2 about here

We have carried out several analyses recently using a large lexical database containing phonetic transcriptions of 126,000 words to study the effects of different constraints on search space reduction in auditory word recognition [55,56]. Figure 2 shows the results of one analysis based on word length constraints. Knowing only the length of a word in phonemes reduces the search space from about 126,000 words to 6,342 words. Clearly, word length is a powerful constraint in reducing the lexicon on the average of about two orders of magnitude, even without any detailed segmental phonetic information. Of course, as Figure 2 shows, the length constraint is strongest for relatively long words.

Figure 2 also shows the results of another analysis, the effect of adding the constraint of minimal phonetic information about every phoneme in the words -- that is, simply classifying each segment as either a consonant or vowel. The reduction in the search space over and above the length constraint by this minimal phonotactic constraint is enormous. The number of words considered as potential candidates for recognition is reduced from the original 126,000 word lexicon to about 34 words per candidate set on average.

Insert Figure 3 about here

Figure 3 shows a comparison of the log weighted-mean candidate set sizes for the minimal phonotactic constraint of classification of phonemes into two categories (consonants and vowels), with the six gross manner class scheme used by Zue and his colleagues. We have found that their results obtained on a 20,000 word lexicon generalize to our 126,000 word lexicon -- the unweighted-mean candidate set size computed in the same way as Shipman and Zue [34] is 2.4 words. Figure 3 also shows the constraint afforded by complete identification of only some of the phonemes in words. While a partial specification of all the phonemes in words provides overall "word shape" information (in some sense), it is not at all clear that complete information about some of the phonemes in a word would be as effective in reducing the search space. We classified just over half of the phonemes in each of the words in the original 126,000 word lexicon from the beginning of the words and from the ends of the words. The results of this analysis are shown in Figure 3. The weighted-mean candidate set size for words classified from the beginning was about 1.7 words; for words classified from the end the weighted mean was about 1.8 words. These results demonstrate that detailed phonetic information, even only partial information about the phonetic content of a word is a very effective heuristic for reducing the number of possible words to be recognized.

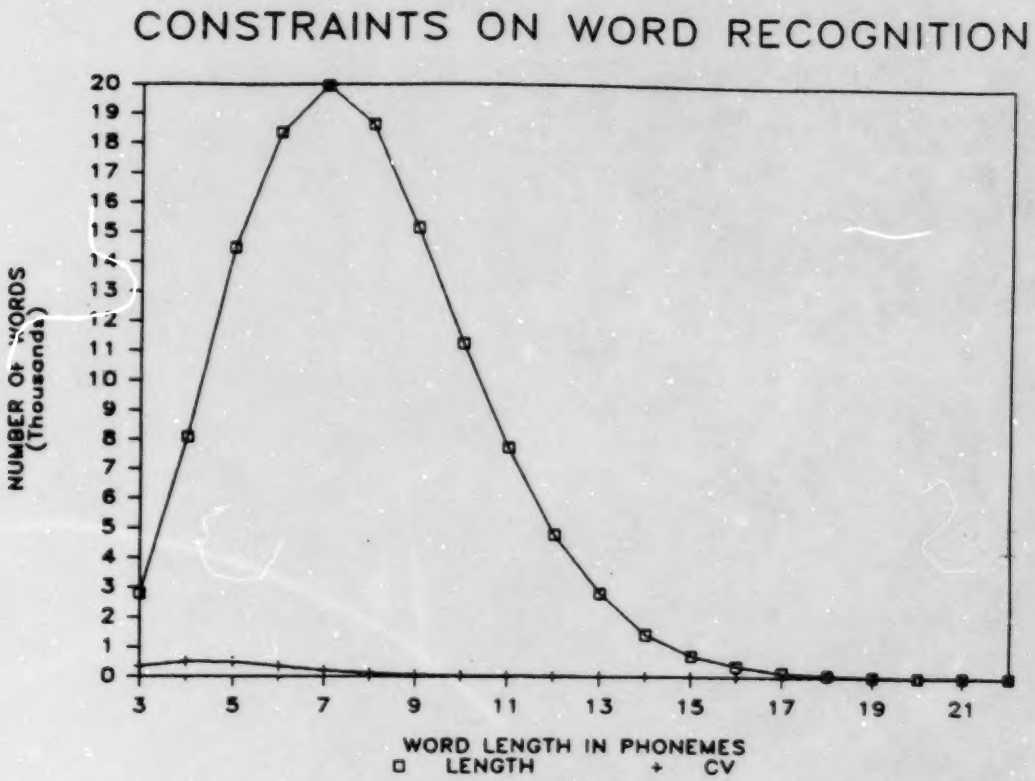

Figure 2. The effects of word length constraints on the number of candidates in a lexicon of 126,000 words. The open squares show the number of words in the database at each word length (in number of phonemes). The plus symbols (+) indicate the increased constraint over and above word length that occurs when each segment is classified as either a consonant or a vowel.

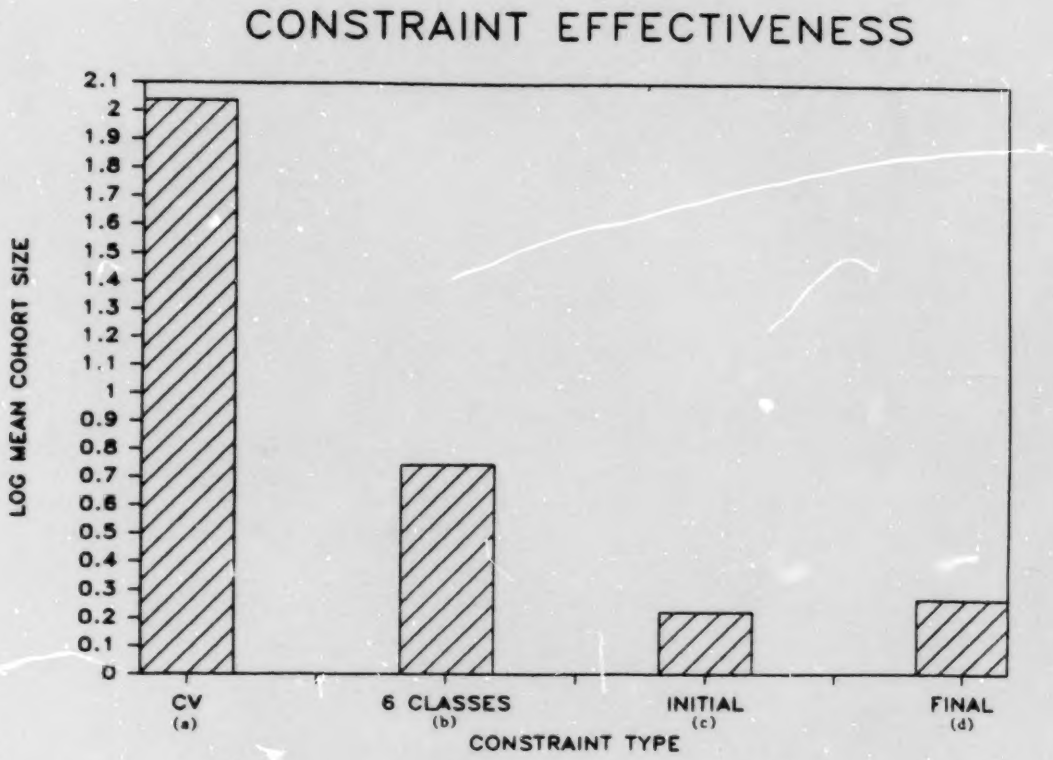

Figure 3. The relative effectiveness of different types of phonetic and phonotactic constraints in a lexicon of 126,000 words. The constraints shown are: (a) every segment in each word classified as a consonant or vowel (CV), (b) every segment labeled as a member of one of six gross manner classes (6 CLASSES), (c) the phonemes in the first half of each word identified exactly (INITIAL), and (d) the phonemes in the last half of each word identified exactly (FINAL). Constraint effectiveness is indexed by the log weighted-mean of the number of word candidates that result from the application of each constraint to the total vocabulary.

Thus, from these initial findings, it is quite apparent that the basic approach of Phonetic Refinement Theory is valid and has much to offer. Having refined roughly the first half of a word, a listener need only compute a fairly coarse characterization of the remainder of a word to uniquely identify it. This finding suggests an account of the observed failure of listeners to detect word-final mispronunciations with the same accuracy as word-initial mispronunciations (see [22]). Given reliable information in the early portions of words, listeners do not need to precisely identify all the phonemes in the latter half of words. Furthermore, the word candidate set sizes resulting from phonetic refinement of the endings of words indicates that, in spite of the large number of common English word-final inflections and affixes, word-final phonetic information also provides strong constraints on word recognition. For words between three and ten phonemes long, the mean cohort size resulting from classification of the beginning was 2.4 words whereas for classifcation of word endings, the mean cohort size was 2.8 words. This small difference in effectiveness between word-initial and word-final constraints suggests that listeners might be slightly better in identifying words from their beginnings than from their endings -- a result that was observed recently with human listeners by Salasoo and Pisoni [28] using the gating paradigm.

Taken together, Phonetic Refinement Theory is able to account for many of the findings in auditory word recognition by reference to structural constraints in the lexicon. Moreover, it is also clear that there are advantages to the phonetic refinement approach with respect to recovery from phonetic classification errors due to noise, ambiguity, or mispronunciations. Phonetic Refinement Theory can account for the ability of listeners to identify words from word endings and their sensitivity to acoustic-phonetic overlap between prime and test words in word recognition. Both of these findings would be difficult, if not impossible, to account for with the current version of Cohort Theory [59] which emphasizes the primacy of word-initial acoustic-phonetic information in controlling activation of potential word candidates in the early stages of word recognition (cf. [29]).

Summary and Conclusions

In this report we have briefly summarized research findings from three on-going projects that are concerned with the general problem of auditory word recognition. Data on the perceptual sensitivity of listeners to the distribution of acoustic-phonetic information in the structure of words, taken together with new research on the organization of words in the lexicon has identified a number of serious problems with the Cohort Theory of auditory word recognition. To deal with these problems, we have proposed a new approach to word recognition known as Phonetic Refinement Theory. The theory was designed to account for the way listeners use detailed knowledge about the internal structure of words and the organization of words in the lexicon in word recognition. Whether the details of our approach will continue to be supported by subsequent research remains to be seen. Regardless of this outcome, however, we feel that the most important contribution of the theory will probably lie in directing research efforts towards the study of how the perceptual processing of the acoustic-phonetic structure of speech interacts with the listener's knowledge of the structure of words and the organization of words in his/her lexicon. This is an important and seemingly neglected area of research on language processing that encompasses both speech perception and word recognition.

References

[1] W. C. Bagley, "The apperception of the spoken sentence: A study in the psychology of language", Am. J. Psych., Vol. 12, 1900-1901, pp. 80-130.

[2] R. A. Cole and A. I. Rudnicky, "What's new in speech perception? The research and ideas of William Chandler Bagley", Psych. Rev., Vol. 90, 1983, pp.94-101.

[3] A. M. Liberman, "On finding that speech is special", Am. Psych., Vol. 37, 1982, pp. 148-167.

[4] B. H. Repp, "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception", Psych. Bull., Vol. 92, 1982, pp. 81-110.

[5] D. B. Pisoni, "Acoustic-phonetic representations in the mental lexicon", Cog, 1985, in press.

[6] D. H. Klatt, "Speech perception: A model of acoustic-phonetic analysis and lexical access", J. Phon., Vol. 7, 1979, pp. 279-312.

[7] W. A. Wickelgren, "Phonetic coding and serial order", in: E. C. Carterette and M. P. Friedman, eds., Handbook of Perception, Vol. VII, Academic Press, 1976, pp. 227-264.

[8] B. T. Lowerre and D. R. Reddy, "The HARPY speech understanding system", in: W. A. Lea, ed., Trends in Speech Recognition, Prentice-Hall, Englewood Cliffs, NJ, 1979.

[9] K. I. Forster, "Accessing the mental lexicon", in: R. J. Wales and E. Walker, eds., New Approaches to Language Mechanisms. North-Holland, Amsterdam, 1976, pp. 257-287.

[10] K. I. Forster, "Levels of processing and the structure of the language processor", in: W. E. Cooper and E. C. T. Walker, eds., Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett, Erlbaum, Hillsdale, N. J., 1979, pp. 27-86.

[11] D. A. Swinney, "The structure and time-course of information interaction during speech comprehension: Lexical segmentation, access, and interpretation", in: J. Mehler, E. C. T. Walker, and M. Garrett, eds., Perspectives on Mental Representations, Erlbaum, Hillsdale, N. J., 1982, pp. 151-167.

[12] D. Norris, "Autonomous processes in comprehension: A reply to Marslen-Wilson and Tyler", Cog., Vol. 11, 1982, pp. 97-101.

[13] J. Morton, "Interaction of information in word recognition", Psych. Rev., Vol. 76, 1969, pp. 165-178.

[14] J. Morton, "Facilitation in word recognition: Experiments causing change in the logogen model", in: P. A. Kolers, M. E. Wrolstad, & H. Bouma, Processing Visible Language, Plenum Press, New York, 1979, pp. 259-268.

[15] J. Morton, "Word Recognition", in: J. Morton and J. D. Marshall, eds., Psycholinguistics 2: Structures and processes, Cambridge: M.I.T. Press, 1979, pp. 107-156.

[16] W. D. Marslen-Wilson and L. K. Tyler, "The temporal structure of spoken language understanding", Cog., Vol. 8, 1980, pp. 1-71.

[17] L. K. Tyler and W. D. Marslen-Wilson, "Conjectures and refutations: A reply to Norris," Cog., Vol. 11, 1982, pp. 103-107.

[18] L. K. Tyler and W. E. Marslen-Wilson, "Speech comprehension processes", in: J. Mehler, E. C. T. Walker, and M. Garrett, eds., Perspectives on Mental Representation, Erlbaum, Hillsdale, N. J., 1982, pp. 169-184.

[19] W. D. Marslen-Wilson and A. Welsh, "Processing interactions and lexical access during word recognition in continuous speech", Cog. Psych., Vol. 10, 1978, pp. 29-63.

[20] R. A. Cole and J. Jakimik, "Understanding speech: How words are heard", in: G. Underwood, ed., Strategies of Information Processing, Academic Press, New York, 1978, pp. 67-116.

[21] D. J. Foss and M. A. Blank, "Identifying the speech codes", Cog. Psych., Vol. 12, 1980, pp. 1-31.

[22] R. A. Cole and J. Jakimik, "A model of speech perception", in: R. Cole, ed., Perception and Production of Fluent Speech, Erlbaum, Hillsdale, N. J., 1980, pp. 133-163.

[23] D. B. Pisoni, "In defense of segmental representations in speech processing", J. Acoust. Soc. Am., Vol. 69, 1981, p. S32.

[24] M. A. Blank, and D. J. Foss, "Semantic facilitation and lexical access during sentence processing", Memory & Cognition, Vol. 6, 1978, pp. 644-652.

[25] R. A. Cole and J. Jakimik and W. E. Cooper, "Perceptibility of phonetic features in fluent speech", J. Acoust. Soc. America, Vol. 64, 1978, pp. 44-56.

[26] J. Morton and J. Long, "Effect of word transitional probability on phoneme identification", J. Verbal Learn. Verbal Behav., Vol. 15, 1976, pp. 43-52.

[27] F. Grosjean, "Spoken word recognition and the gating paradigm", Percept. Psychoph., Vol. 28, 1980, pp. 267-283.

[28] A. Salasoo and D. B. Pisoni, "Sources of knowledge in spoken word recognition", J. Verbal Learn. Verbal Behav., Vol. 0, 1984, pp. 000-000.

[29] H. C. Nusbaum and L. M. Slowiaczek, "An activation model of the Cohort theory of auditory word recognition", Sixteenth Annual Meeting of the Society for Mathematical Psychology, Boulder, Colorado, August, 1983.

[30] L. M. Slowiaczek and D. B. Pisoni, "Acoustic-phonetic priming in auditory word recognition: Some tests of the Cohort theory", Research on Speech Perception, Progress Report No. 8, Speech Research Laboratory Indiana University, Bloomington, 1982, pp. 3-25.

[31] L. M. Slowiaczek, H. C. Nusbaum and D. B. Pisoni, "Acoustic-phonetic priming in auditory word recognition", Cognitive Psychology, 1984, submitted.

[32] D. E. Meyer, R. W., Schvaneveldt and M. G. Ruddy, "Loci of context effects in visual word recognition", in: P. M. A. Rabbitt and S. Dornic, eds., Attention and Performance V, Academic Press, New York, 1975.

[33] M. K. Tannenhaus, H. P. Flanagan, and M. S. Seidenberg, "Orthographic and phonological activation in auditory and visual word recognition", Memory and Cognition, Vol. 8, 1980, pp. 513-520.

[34] D. W. Shipman and V. W. Zue, "Properties of large lexicons: Implications for advanced isolated word recognition systems", Proceedings of the 1982 IEEE International Conference on Acoustics, Speech and Signal Processing, Paris, France, April 1982.

[35] T. K. Landauer and L. A. Streeter, "Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition", J. Verbal Learn. Verbal Behav., Vol. 12, 1973, pp. 119-131.

[36] B. Eukel, "A phonotactic basis for word frequency effects: Implications for automatic speech recognition", J. Acoust. Soc. America, Vol. 68, 1980, p. S33.

[37] J. H. Greenberg and J. J. Jenkins, "Studies in the psychological correlates of the sound system of American English", Word, Vol. 20, 1964, pp. 157-177.

[38] Kenyon and Knott, A Pronouncing Dictionary of American English Springfield, MA: G. C. Merriam, 1953.

[39] Webster's Seventh Collegiate Dictionary, Los Angeles: Library Reproduction Service, 1967.

[40] D. Scarborough, C. Cortese, and H. Scarborough, "Frequency and repetition effects in lexical memory", J. Exp. Psych.: Hum. Percept. Perform., Vol. 3, 1977, pp. 1-17.

[41] D. E. Broadbent, "Word-frequency effect and response bias", Psych. Rev., Vol. 74, 1967, pp. 1-15.

[42] C. R. Brown and H. Rubenstein, "Test of response bias explanation of word-frequency effect", Science, Vol. 133, 1961, pp. 280-281.

[43] J. Catlin, "On the word-frequency effect", Psych. Rev., Vol. 76, 1969, pp. 504-506.

[44] M. Triesman, "On the word frequency effect: Comments on the papers by J. Catlin and L. H. Nakatani", Psych. Rev., Vol. 78, 1971, pp. 420-425.

[45] M. Triesman, "Space or lexicon? The word frequency effect and the error response frequency effect", J. Verbal Learn. Verbal Behav., Vol. 17, 1978, pp. 37-59.

[46] P. B. Denes, "On the statistics of spoken English", J. Acoust. Soc. America, Vol. 35, 1963, pp. 892-904.

[47] P. A. Luce, "Structural distinctions between high and low frequency words in auditory word recognition", Unpublished doctoral dissertation, Indiana University, 1985.

[48] P. A. Luce and D. B. Pisoni, "Speech perception: Recent trends in research, theory, and applications," to appear in: H. Winitz, ed., Human Communication and Its Disorders, Ablex, Norwood, N. J., 1984.

[49] F. Kucera and W. Francis, Computational Analysis of Present Day American English, Brown University Press, Providence, R. I., 1967.

[50] G. A. Miller, Language and Communication, McGraw Hill, New York, 1951.

[51] M. Chen and W. S-Y. Wang, "Sound change: Actuation and implementation", Lang., Vol. 51, 1975, pp. 255-281.

[52] J. D. Hood and J. F. Poole, "Influence of the speaker and other factors affecting speech intelligibility", Audiology, Vol. 19, 1980, pp. 434-455.

[53] D. C. Anderson, "The number and nature of alternatives as an index of intellibility", Unpublished doctoral dissertation, Ohio State University, 1962.

[54] L. L. Havens and W. E. Foote, "The effect of competition on visual duration threshold and its independence of stimulus frequency", J. Exp. Psych., Vol. 65, 1963, pp. 6-11.

[55] H. C. Nusbaum and D. B. Pisoni, "Human speech perception: Implications for computer speech recognition", in: W. A. Lea, ed., Towards Robustness in Speech Recognition, 1984, forthcoming.

[56] H. C. Nusbaum and D. B. Pisoni, "Phonetic refinement: Auditory word recognition by constraint satisfaction", Manuscript in preparation, 1984.

[57] S. W. Zucker, "Vertical and horizontal processes in low level vision", in: A. R. Hanson and E. M. Riseman, eds., Computer vision systems, New York: Academic Press, 1978.

[58] L. K. Tyler and J. Wessels, "Quantifying contextual contributions to word-recognition processes", Percept. & Psychoph., Vol. 34, 1983, pp. 409-420.

[59] W. D. Marslen-Wilson, "Function and process in spoken word recognition: A Tutorial Review", in: H. Bouma & D. G. Bouwhuis, eds., Attention and Performance X. Control of Language Processes, Hillsdale, NJ: Lawrence Erlbaum, 1984, pp. 125-150.

[60] S. G. Nooteboom, "Lexical retrieval from fragments of spoken words: Beginnings vs. endings", J. Phon., Vol. 9, 1981, pp. 407-424.

[61] K. E. Stanovich and R. F. West, "On priming by a sentence context", J. Exp. Psych.: Gen., Vol. 112, pp. 1-36.

[62] P. A. Luce, "Context and frequency effects in spoken word recognition", 1st Conference on Hoosier Mental Life, West Lafayette, IN, May 1983.

[63] T. H. Crystal, M. K. Hoffman and A. S. House, "Statistics of phonetic category representation of speech for application to word recognition", Institute for Defense Analyses, Princeton, NJ, September 1977.

[64] D. P. Huttenlocher and V. W. Zue, "A model of lexical access based on partial phonetic information", Proceedings of ICASSP-84, Vol. 2, 1984.

[65] A. Waibel, "Suprasegmentals in very large vocabulary isolated word recognition", Proceedings of ICASSP-84, Vol. 2, 1984.